MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 970
C.B.I.P. Memo No. 027

October 1987

VISUAL INTEGRATION AND DETECTION OF DISCONTINUITIES: THE KEY ROLE OF INTENSITY EDGES

Ed Gamble and Tomaso Poggio

Abstract: Integration of several vision modules is likely to be one of the keys to the power and robustness of the human visual system. The problem of integrating early vision cues is also emerging as a central problem in current computer vision research. In this paper we suggest that integration is best performed at the location of discontinuities in early processes, such as discontinuities in image brightness, depth, motion, texture and color. Coupled Markov Random Field models, based on Bayes estimation techniques, can be used to combine vision modalities with their discontinuities. These models generate algorithms that map naturally onto parallel fine-grained architectures such as the Connection Machine. We derive a scheme to integrate intensity edges with stereo depth and motion field information and show results on synthetic and natural images. The use of intensity edges to integrate other visual cues and to help discover discontinuities emerges as a general and powerful principle.

Acknowledgments. This report describes research done within the Artificial Intelligence Laboratory. Support for the A.I. Laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research (ONR) contract N00014–85–K–0124. Support for this research is also provided by a grant from ONR, Engineering Psychology Division and by a Hughes Aircraft Corporation gift to the Artificial Intelligence Center for T. Poggio.

1 Introduction

One of the keys to the reliability, flexibility and robustness of biological visual systems is their ability to integrate several different visual cues. Early vision processes such as stereo, motion, texture, shading and color give separate cues to the distance of three-dimensional surfaces from the viewer and to their material properties. Integration of the evidence provided separately by these cues can provide a more reliable map of the surfaces and their properties than any single cue alone.

Thus visual integration is likely to be a key to understanding biological visual systems and to developing robust vision machines. Existing methods do not seem capable of providing a general solution. Standard regularization[2] provides a common framework for many early vision problems and leads to the minimization of quadratic energy functionals. If standard regularization is used to integrate information from different processes, the energy functional consists of the sum of quadratic parts, each associated with a separate process. This implies that the result is a linear combination of the different cues (possibly with space-varying coefficients). Linear combination – say of depth from stereo and from shading – does not seem, however, a flexible enough integration method. Even more important, no instances of standard regularization can handle discontinuities, because the solution space is restricted to generalized splines[21,2]. As we will explain later, we believe that detecting and representing discontinuities (for instance depth discontinuities) is a key part of the integration step[21].

To overcome these difficulties we have developed an extension of regularization that promises to deal simultaneously with discontinuities and with the integration of vision modules. This extension is based on the use of coupled Markov Random Fields[1], introduced recently by Geman and Geman[9] and extended by Marroquin, Mitter and Poggio[19]. The standard regularization method for vision is a special case of this new approach.

1.1 The Role of Discontinuities

One of the most important constraints for recovering surface properties is that the physical processes underlying image formation are typically smooth:

[1] A different, interesting approach has be explored by Blake[3]

depth and orientation of surfaces are *mostly* continuous and so are reflectance and illumination. The smoothness property is captured well by standard regularization. Surfaces and their properties, however, are not always smooth: they are smooth *almost* everywhere, but not at discontinuities. Lines of discontinuity are themselves usually continuous, relatively smooth, nonintersecting curves. It is critical to detect the discontinuities reliably, because they usually represent the most important locations in a scene: depth discontinuities, for instance, often correspond to the boundaries of an object or of a part. Furthermore, discontinuities play a critical role in fusing information from different physical processes. The reason is clear: in smooth regions, the physical processes are coupled together by the imaging equation, and all contribute to image formation. However, the coupling is difficult to know precisely: it depends on quantities such as the form of the reflectance function. The effects of discontinuities are instead robust and qualitative: for instance, depth discontinuities usually correspond to intensity edges. Therefore, discontinuities are ideal places for integrating information. Furthermore, *partial* information about discontinuities in a single process can be detected relatively easily. Several types of motion discontinuities, for example, can be measured with simple operations on the time-dependent intensity array, especially if the interframe interval is small. Partial albedo discontinuities also are often detectable using simple operations. Intensity edges are detected quite reliably by the Canny edge detector. However, the fast, rough detection of discontinuities performed by these early operations is noisy and incomplete: it must be refined by integrating them across processes and by exploiting constraints on the continuity of discontinuities.

In summary, discontinuities: 1) represent the most useful information, 2) are easy to detect (though in a partial and possibly noisy way) and 3) provide good locations to integrate different cues.

1.2 Coupled Markov Random Fields

Markov Random Fields for image modeling have seen increasing use since the work of Geman and Geman[9]. Their utility for image modeling derives from several MRF characteristics. MRFs provide a natural way to impose general image properties of smoothness and continuity, for example of depth and motion, while also incorporating discontinuities. Bayes' rule establishes a relationship between the possibly corrupted observed data and

the desired scene data. Solution methods are available, though often time consuming. Some recent MRF applications have involved scene segmentation using depths[18], texture[6] and motion[20].

A Markov Random Field on a lattice can be represented as a lattice of sites, each one with a random variable. The value depends probabilistically on the value of neighboring sites. The rules governing this local dependence can be given in a variety of ways and can be made to capture constraints such as the continuity of a surface (if the MRF represents depth values).

Our idea is to associate a MRF on a lattice to each physical process to be integrated and another (binary) MRF to its discontinuities (see figure 1). The lattices are coupled to each other to reflect the interdependence of the corresponding processes in image formation. Thus the various MRFs mirror the different physical events that underlie image formation: surface and surface discontinuities, spectral albedo and albedo discontinuities, shadows, surface normal, and so on. Physical constraints apply to each of these processes independently. In addition, there are constraints between these processes (for instance between depth and surface normal). The image data constrain the way the processes combine. Note that consideration of sequences of images in time will introduce additional powerful constraints such as rigidity. The constraints on the surfaces are local conditions (such as smoothness, necessary mainly because of its regularizing role in the face of omnipresent noise) valid everywhere *except* at discontinuities. As we discussed earlier, discontinuities are critically important and should be detected early.

Notice that the coupling of the line process with the associated continuous process provides a module that *combines region-based with boundary-based segmentation* (see figure 1).

The local potentials underlying the *a priori* probability distribution of the MRFs represent the constraints on the physical processes (smoothness, positivity, values within certain bounds, etc.); the coupling between MRFs represents the compatibility constraints *between* processes. The device of coupled MRFs provides an ideal tool to impose local constraints such as smoothness, allowing at the same time an explicit role for discontinuities through the *line processes*[9] and similar processes such as *occlusions*[19]. Our new idea is to incorporate additional *observable* discontinuity data provided by algorithms specialized to detect sharp changes in the observed properties of intensity, motion, stereo disparity, texture, and so on. The observable discontinuities

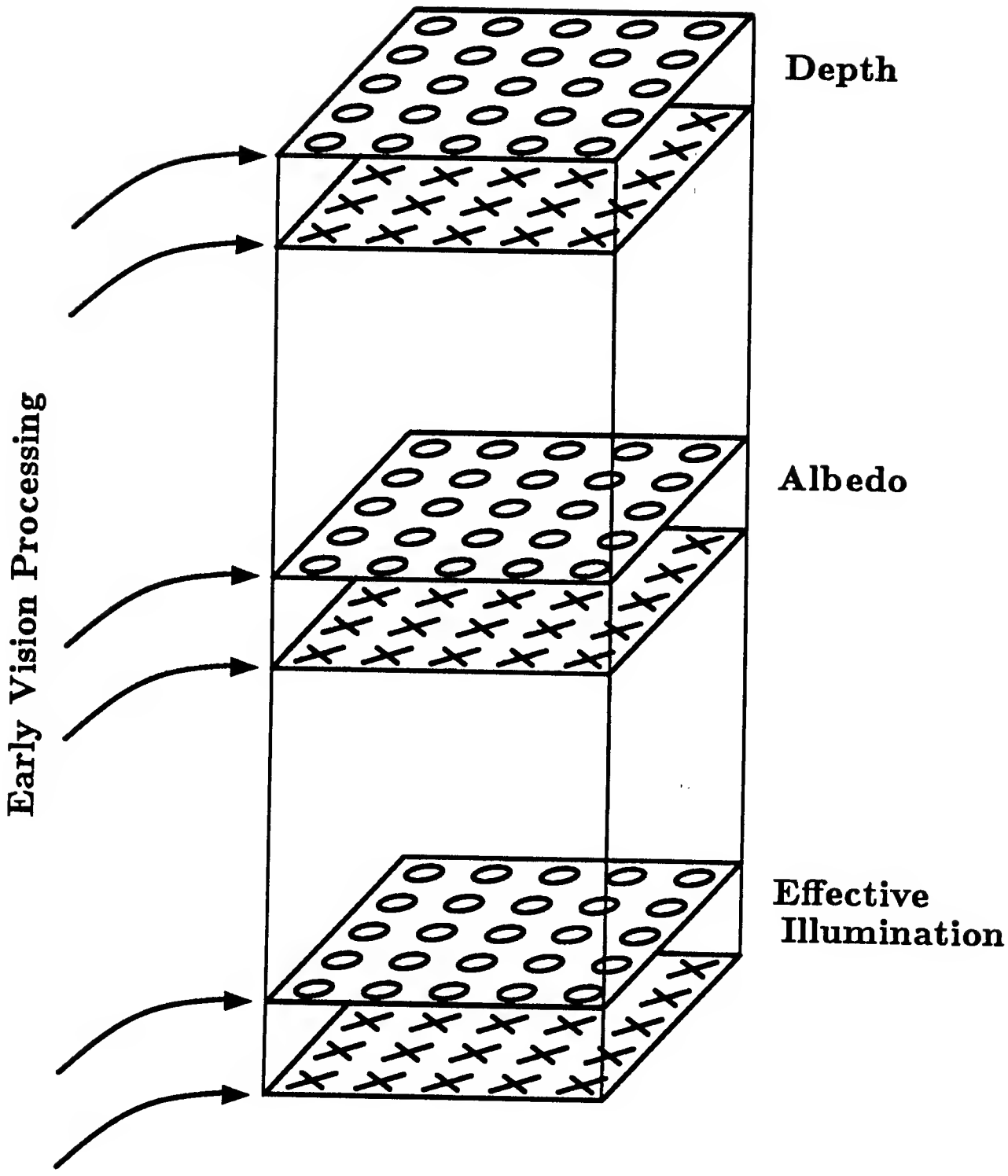

Figure 1: MRF lattices representing the output of different early processes and their discontinuities (the crosses represent the sites of the binary line processes). Each representation, for instance depth, is coupled to its discontinuities and to other cues such as intensity or motion.

provide an initial rough solution to the segmentation problem. Using the MRFs for estimating the fields gives increasingly precise solutions, simultaneously *filling in* the continuous regions that are only sparsely observable. The solution at each iteration is available to later modules, such as recognition.

1.3 The Key Role of Intensity Edges

One of the results of our integration work is that intensity edges play primary role in guiding the search for discontinuities in other processes (for instance depth). The point seems so important that we would like to phrase it as a rather general conjecture on the proper organization of the integration stage: *intensity edges guide the detection of discontinuities in the other physical processes, thereby coupling surface depth, surface orientation, shadows, specularities and surface markings to the image data and to each other.*

The reason for the critical role of intensity edges is intuitively clear – usually changes in surface properties (depth, orientation, material, texture) produce large intensity gradients in the image. Under the assumption of opacity and of a simple imaging model (the reflectance function is assumed to contain a lambertian and a specular term), there are six physical causes for large intensity gradients in the image: occluding edges (*extremal edges* and *blades*), folds, shadow edges, surface markings and specular edges. In addition, motion discontinuities are usually coupled to intensity edges. It is for exactly this reason that edge detection is so important in artificial – and probably also biological – vision.

1.4 Plan of the Paper

In this paper we introduce a method for detecting and reconstructing depth discontinuities by using the information provided by intensity edges. We do the same for motion discontinuities. First we introduce the Markov Random Field formalism. The use of intensity edges for surface interpolation is discussed next, together with the derivation of the associated MRF model. We then describe our Connection Machine implementation and the results on synthetic and real data. Finally the discussion focuses on the open problems and on the implications of our results for the general problem of integrating all vision modules.

2 Coupling Intensity Edges with Sparse Depth Data

To illustrate our approach we consider the specific and important problem of computing an approximate surface and especially the surface depth discontinuities from sparse depth data[10,25,18]. The main new idea here is to exploit the integration of additional vision cues. In particular we describe a scheme in which intensity edges are integrated with sparse depth data. Sparse depth data arise from the output of feature–based stereo algorithms. Typical stereo algorithms provide depth data at a subset of image features[15,10,8]. These features might be a Laplacian filter's zero-crossings from one of the intensity images. The depth information is computed by measuring pixel displacements (disparity) between corresponding image features. As is typical of all known stereo algorithms, the disparities are plagued by errors precisely at depth discontinuities where surfaces are usually occluded.

The problem, then, is to smooth and fill in the sparse depth data (i.e., reconstruct the surface), while detecting the critically important depth discontinuities. Prior attempts at depth discontinuity identification allowed the discontinuities to form anywhere in the image provided the depth difference between neighboring sites was significant[18,24]. Due to the sparseness and noise in the depth data, the identified discontinuities are: 1) offset from and 2) ragged or wiggly compared with the *correct* discontinuities. These limitations become more serious when the images contain a large range of depth differences, as in natural images.

Because of the constraints on image formation discussed earlier, the *correct* depth discontinuities will, in almost all cases, correspond precisely to the locations of intensity edges. Our integration scheme exploits this by restricting depth discontinuity formation to a subset of the intensity edges. This restriction ensures that the smoothness and continuity of discontinuities can be no worse than the intensity edges themselves. In addition, the difficult problem of MRF parameter specification is simplified since this integration scheme proves less sensitive to MRF parameter variations, particularly when the depth data contain a large range of depth differences.

There are some cases in which discontinuities will not occur at intensity edges. Any object that *blends in* with its background presents such a case. This situation occurs rarely in natural scenes; yet, for practical reasons such

as camera underexposure or saturation, the object may blend in with the background at some locations. However, for these cases, the point is somewhat moot, since without intensity edges, feature-based stereo or motion algorithms will not provide depth or motion data.

A more general situation arises when the features used for stereo or motion are different from the discontinuity–limiting features. This is desirable since the continuity constraints used by stereo and motion algorithms assume that the features used for matching are located on surfaces. Thus stereo and motion algorithms should use high resolution, dense features that identify surface markings as opposed to bounding contours which in general correspond to surface locations that are different in the two images of a stereo pair. The discontinuity–limiting features however can be chosen to better correspond to object boundaries.

The results section contains examples in which the discontinuities are identified and the surface reconstructed both with and without the benefit of intensity edge information. The next section presents a limited overview of MRF particulars and contains the appropriate MRF energy function for integrating intensity edges with, in this case, the sparse depth data produced by a stereo algorithm.

3 MRF Formulation for Stereo and Intensity Edge Coupling

The theory of Markov Random Fields can be found elsewhere[9,17]. We present only an overview here followed by a description of the energy functions used for integration.

The Hammersley-Clifford theorem states the equivalence between a MRF and a Gibbs distribution as follows. If X is a MRF on a lattice S with respect to the neighborhood system G, then $P(X = \omega)$ is given by:

$$P(X = \omega) = \frac{1}{Z} e^{-\frac{1}{T}U(X)} \tag{1}$$

Z is a normalization factor, T is the *temperature* and $U(X)$ is the *energy function.* The temperature parameter, T, could be absorbed into $U(X)$; however, when the solution method is discussed, T proves useful as a separate

variable. The energy function is of the form:

$$U(X) = \sum_C U_C(X). \tag{2}$$

The sum of the *potentials*, $U_C(X)$, is over the neighborhood's *cliques*. A clique is either a single lattice site or a set of lattice sites such that any two sites belonging to it are neighbors of one another. The function $P(X = \omega)$ is called the *prior* distribution and abbreviated here by $P(X)$.

The prior distribution on X, where X, for example, might be the reconstructed surface, must be determined based on some observations or input data, Y. To relate X to Y Bayes' formula is used,

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}. \tag{3}$$

The observations, Y, are obtained conceptually by degrading X, such as by the addition of noise or blurring. If the type of degradation is known, the distribution $P(Y|X)$, can be computed. Marroquin[17] has shown that for the case of zero-mean white Gaussian noise, $P(Y|X)$ is a Gibbs distribution with potential:

$$U(Y|X) = \sum_{i \in S} U_i(Y|X); \quad U_i(Y|X) = -\alpha\gamma_i(x_i - y_i)^2. \tag{4}$$

The sum is over all lattice sites and

$$\gamma_i = \begin{cases} 1, & \text{if input data exists at lattice site } i \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

When this result for $P(Y|X)$ is combined with the MRF prior distribution, $P(X)$, and Bayes' rule the *a posteriori* distribution $P(X|Y)$ is:

$$P(X|Y) = \frac{1}{Z} \exp\left\{-\frac{1}{T}\sum_i U_i(X|Y)\right\} \tag{6}$$

for $U_i(X|Y) = U_i(X) + U_i(Y|X)$ and with Z a normalization constant independent of X. This *a posteriori* distribution provides the likelihoods for all possible states X, given the observable data Y.

Given the posterior distribution $P(X|Y)$ and the *external field* Y the desired field X can be retrieved once a suitable error criterion is specified. The

Maximizer of the Posterior Mean (MPM) reduces the problem of annealing and has been successfully applied for our results. With the criterion specified, the relaxation algorithm for solution is largely determined. The question of a suitable error criterion and algorithmic consequences has been thoroughly discussed by Marroquin[17].

The problem has now become one of specifying the MRF potentials, $U_i(X)$ and $U_i(Y|X)$. The potentials impose the physical constraints of continuity and smoothness of surfaces (except at depth discontinuities) along with continuity and smoothness of depth discontinuities. These constraints are imposed by tailoring the energy function to minimize the energy (maximize the probability) when the state occupied satisfies the desired physical constraints. Typically this choice is empirical although one might envisage estimating the *prior* associated with, for instance, depth smoothness from a specific class of surface data.

The MRF state space used herein is similar to that of Geman and Geman[9] along with Marroquin[17] where each lattice site is composed of a depth process and two line processes, $X = \{F, L\}$. The depth process, F, is a continuous random variable whose value is related to the distance of a surface point from the observer. The value of F at site i is denoted as f_i where $-\infty \leq f_i \leq \infty$. The depth process neighborhood system to site i consists of the four nearest neighbors: east, south, west and north, to i. Although a continuous random variable should not be updated using the Heat Bath algorithm, the depth process can be deterministically updated[17], provided the MRF energy is suitably defined. Figure 2 illustrates the MRF lattice with the depth and line processes.

The line process used here, L, contains a vertical and horizontal orientation that are conceptually located between lattice sites. The vertical line process is located between its lattice site and the neighboring eastern lattice site, whereas the horizontal line process separates its lattice site and the nearest southern lattice site. Each orientation is a binary random field, $l_i^j \in \{0, 1\}$ where the scripts on l_i^j denote the line process that separates lattice site i from j. The horizontal line process at site i is denoted as l_i^h; the vertical line process is l_i^v. Smoothing of the depth process is inhibited when the line state is *on*, $l_i^j = 1$, since smoothing should not occur across depth discontinuities; otherwise, depth process smoothing is performed. An *on* state signifies the presence of a depth discontinuity. The conditions for

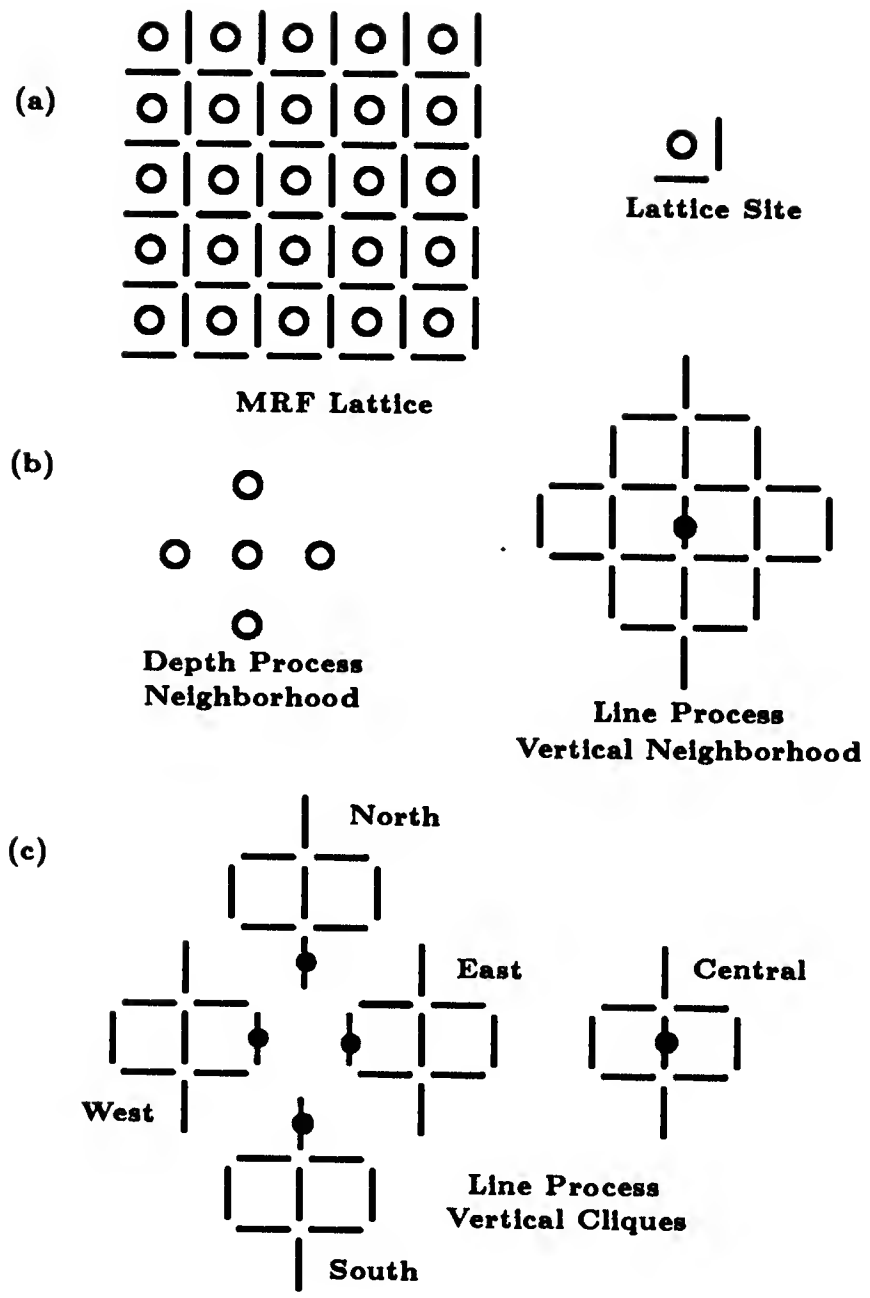

Figure 2: (a) A lattice site is composed of a single depth process (illustrated with a circle) along with a vertical and a horizontal line process. The MRF Lattice consists of a rectangular grid of these lattice sites. (b) The neighborhood for the depth process and the vertical line process neighborhood. The black dot in the line process neighborhood indicates the lattice site for this neighborhood. (c) The five maximal cliques (north, east, south, west and central) for the vertical line process are shown. In this paper we only consider configurations of the central clique. This is equivalent to assigning zero energies to all configurations of the other four cliques.

depth discontinuity formation are encapsulated in the MRF energy function presented subsequently.

The external fields to the MRF are the sparse depth information and the intensity edges. The sparse depths, G, are represented by two variables, g_i and γ_i for site i. The value g_i is analogous to f_i; it is continuously valued over the real numbers, although in practice, since g_i is provided by stereo output, it is discrete. The variable γ_i encodes the sparseness of the stereo output and is defined as in equation 5.

The intensity edges are represented by the field, E. This field is similar to the line process, L, except that $e_i^j = 1$, rather than indicating the presence of a depth discontinuity, *permits the formation* of a depth discontinuity between lattice site i and neighbor j. The MRF energy is designed so that $e_i^j = 0$ implies (in the present implementation) $l_i^j = 0$ for all $i, j \in S$. An edge detector, such as Canny's[4], will mark *a site* i as an edge, but e_i^j marks potential discontinuities *between sites* i and j. To resolve this ambiguity, if an edge is at site i, then $e_i^k = 1$ where k is each of the nearest neighbors to site i. This intensity edge field, E, along with G comprise the MRF external field Y such that $Y = \{G, E\}$.

Given the external fields, Y, and the random variables, X, equation 6 provides the posterior distribution with the MRF energy given as

$$U(x|y) = \sum_i U_i(x|y)$$

$$U_i(x|y) = \alpha\gamma_i(f_i - g_i)^2 + \sum_{j \in nn}(1 - l_i^j)(f_i - f_j)^2 + \sum_{j \in <h,v>} \left[\beta U_C(l_i^j) + \beta'(1 - e_i^j)l_i^j\right]. \quad (7)$$

The first term in this equation is the coupling between the depth process and the sparse and noisy input data. The coupling factor, α, is related to the noise in g. For noiseless data, $\alpha \to \infty$ thereby ensuring $f_i = g_i$. Otherwise, when $\alpha = 0$ no input data coupling occurs and f is smoothed by the term involving $(f_i - f_j)^2$ in equation 7. The precise relation between α and the noise depends on the noise model assumed. For a model of measurement that includes Gaussian random noise

$$\alpha = \frac{1}{\sigma^2}$$

where σ is the gaussian's half width at half maximum[17]. Note that if the noise model's parameters vary locally, it might be appropriate to vary α locally as

$$\alpha_i = \frac{1}{\sigma_i^2}.$$

Local variation in noise parameters does occur in the stereo algorithm of Drumheller and Poggio[7]; this variation is reflected in the stereo match scores of that algorithm. The present paper does not address this issue; here we keep α constant, usually in the range 0.1 to 2.0. The input data coupling to f occurs when $\gamma = 1$. Typically 5 to 10% of the lattice sites have input depths associated with them.

The last term in equation 7 implements the integration scheme between sparse stereo depths and intensity edges. The term forbids depth discontinuity formation except where an external edge exists. Discontinuity formation is prevented by letting $\beta' \to \infty$. When $l_i^j = 1$ and $e_i^j = 0$, this term contributes a large energy, $U_i(x|y) \to \infty$ and the associated probability for $l_i^j = 1$ is zero. At sites where $e_i^j = 1$ this energy term contributes nothing and the depth discontinuity formation is determined by the other factors in equation 7. The problems of misalignment might be handled by suitably modifying this term in the energy $U_i(x|y)$ to produce a it cone of influence or, for a simple case, by "thickening" the input intensity edges. For instance, we may use instead of e_i^j in equation 8, $e_i^j * G$, where $*$ denotes convolution and G is a gaussian or another appropriate cone of influence function. The results presented in this paper do not utilize a cone of influence.

The second and third terms in equation (7) encapsulate our prior expectations concerning depth discontinuities and surface reconstruction. They compose the potential $U(X)$ of the prior distribution (equation 1). These two terms 'compete' in the sense that turning on a line costs energy $\beta U_C(l_i^j)$ but saves energy $(f_i - f_j)^2$. The interplay of these two potentials largely determines the formation of depth discontinuities where $e_i^j = 1$. The second term couples the line and depth processes, the third term determines the line process clique energy. This line and depth process coupling is summed over the nearest neighbors, nn, to site i, with each neighbor contributing an energy $(f_i - f_j)^2$ when $l_i^j = 0$.

The quadratic term, $(f_i - f_j)^2$, tends to smooth the depth process since it is minimized when $f_i = f_j$. Depth discontinuities have a higher probability of forming when the energy to create a line, $\beta U_C(l_i^j)$, is less than this energy

to smooth the depths. The factor β is a free parameter that determines what size depth difference is likely to produce a depth discontinuity. Specification of β is largely image dependent and, although a suitable range has been determined, a general theory specifying β remains elusive. The line process clique energy will be examined in detail later.

The Heat Bath algorithm cannot be simply applied to equation 7 since the f_i are continuous variables. Instead we employ a technique to smooth the depth process deterministically, but to update the line process stochastically with the Heat Bath algorithm[17]. With the line process state fixed, the MRF energy of equation 7 is non-negative definite quadratic with a stable and unique fixed point for the f_i (practically, β' never contributes since the configuration $e_i^j = 0$ and $l_i^j = 1$ has a vanishing probability). In this situation, the depth process can be smoothed deteministically to find the fixed point. After this fixed point in depth is determined, the line process is stochastically updated, the new fixed point in depth is determined and the scheme is repeated.

Once the line process approaches equilibrium (roughly 1000 iterations), statistics are gathered to compute the MPM estimate. The MPM estimate is computed from $P(l_i^j = 1) = \frac{1}{n}\sum l_i^j$, where n is the number of iterations over which statistics are gathered[17]. When $P(l_i^j = 1) \geq (0.5 + 1/\sqrt{n})$, statistical fluctuations about 0.5 are reduced and the MPM estimate is turned on to mark a discontinuity. Use of the MPM estimate does not require annealing but the *a posteriori* distribution's coupling parameters must produce a reasonable amount of line process agitation thereby sampling much of the line process sample space.

3.1 Choice of Line Clique Energies

Figure 2 shows the line process neighborhood for the vertical line process. Of the five cliques shown for this neighborhood, only the clique centered about the vertical lattice site has, by design, a non-zero potential $U_C(l_i^j)$. This potential depends on the 256 possible configurations associated with the clique. The desirable configurations are a small subset of all possible configurations and they impose the constraints of smoothness and continuity on the depth discontinuities. These constraints are embodied in the following five heuristics which divide the desirable configurations into *classes*:

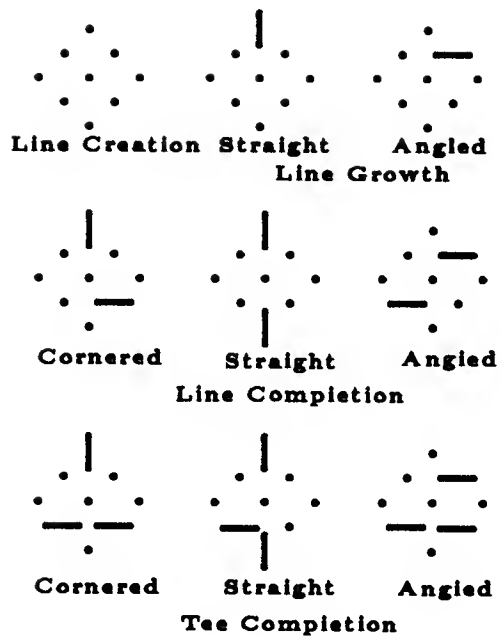

Figure 3: The four classes of non-forbidden line configurations for the vertical line process. A dot, '.' represents an off state; on states are shown with their oriented lines. The symmetry operations producing the other allowed configurations are discussed in the text. The horizontal line process configurations are identical provided the vertical line process cliques are rotated by 90 degrees.

- Turn on a lone site provided a 'large' depth discontinuity is present [*Line Creation*].
- Turn on a site extending an already present line segment even if the depth discontinuity is 'small' [*Line Growth*].
- Always turn on a site if doing so would connect two line segments [*Line Completion*].
- Allow tees to occur infrequently where supported by at least a 'small' depth discontinuity [*Tee Completion*].
- All other configurations should occur rarely if at all [*Forbidden*].

Examples of the first four classes are shown in figure 3. In addition to these configurations, three symmetry operations produce the other non-

forbidden classes. These symmetry operations are: rotation by 180 degrees about an axis perpendicular to the page, reflection about the vertical axis (for the vertical line process orientation) and the 180 degree rotation followed by the reflection operation. With these symmetry operations and clique classes, a total of 22 unique configurations are allowed from the original set of 256. When $l_i^v = 0$ (line is off), the clique potential is 0. However, when $l_i^v = 1$, the clique energy is determined by the five classes; this is the energy required to turn on the line.

The line process clique considered here is only one of the cliques associated with the neighborhood shown in figure 2. In previous work[9,17], the smaller neighborhood did not readily produce lines of any orientation; the cliques tended to create vertical or horizontal line segments. The 'large' neighborhood used here (though incompletely, because we assign zero energies to several cliques), does encourage isotropic line formation without exacting too high a computational penalty.

4 Stereo and Synthetic Image Results

The MRF scheme for coupling intensity edges to sparse stereo depth data has been implemented on a Connection Machine[11]. The sparse depth data and intensity images from both real stereo and synthetic images have been examined. This section presents these image results for some typical images.

4.1 Connection Machine Implementation

The Connection Machine (CM) is a fine-grained parallel computer manufactured by Thinking Machines Corporation. We used their CM-1 model with 16k processors. Each processor is connected to its four nearest neighbors (north, east, south and west) in a two-dimensional grid, the NEWS network, and each 16 processor group is connected to a 12-dimensional hypercube, the Router. These two communication modes allow fast access between neighboring processors and logarithmic-time access between any two processors. Each processor is a simple 1-bit processor with 4 kilobits of memory. All processors execute a single instruction stream. The CM was configured to match the image size, 256 x 256, by using virtual processors.

For the MRF implementation each CM processor represents an MRF lattice site. This configuration proves ideal for implementing the MRF cliques over the CM NEWS network. The limited number of non-forbidden line clique states and energies are stored in tabular form at each processor. Determination of the line clique state requires access to the four nearest neighbors plus the north-east (south-west) neighbor for the vertical (horizontal) orientation. At the image borders, the line processes are always *on*, thereby conveniently preventing depth process smoothing beyond the borders.

The MRF input data was obtained from two previously implemented CM-1 algorithms. For the real stereo depth data, MIT's Eye-Head system provided the stereo pair and the Drumheller-Poggio CM-1 stereo algorithm[8] produced the disparity data at a subset of DOG zero-crossing features. The intensity edges came from Todd Cass' [13] implementation of Canny's edge detector. These edges do not coincide with the stereo algorithm features.

When synthetic data was used, the image depths were produced by the TMC 3-D Toolkit as was a dense depth map. A sparse map was obtained by randomly discarding 90 to 95 percent of the depth values. Uniformly distributed random noise was added to the synthetic sparse depth data.

The initial line process state is set to mimic the intensity edge map as provided by the Canny edge detection stage. The MRF depth values are created by using the sparse input depths to "brush fire fill" and then by deterministically smoothing the depth values. During the deterministic smoothing of the initial depth process, the depth external field coupling, α, is infinite.

4.2 Results

Figure 4 shows the MRF results on a synthetic image for two intensity edge coupling schemes. In the first scheme, intensity edges are not used in the MRF process. This allows depth discontinuities to form anywhere and is achieved by setting $e_i^j = 1$ for all $i, j \in S$. The upper left image shows the synthetic scene from which the sparse depth data was derived. The lower left image in Figure 4 illustrates the depth discontinuities identified with the MPM estimate of the MRF process. When the depths vary rapidly, many closely spaced discontinuities are formed. These discontinuities are ragged and also displaced from the actual object boundaries (as marked by intensity edges). The reconstructed depth surface is not shown.

The second scheme strongly penalizes depth discontinuity formation ev-

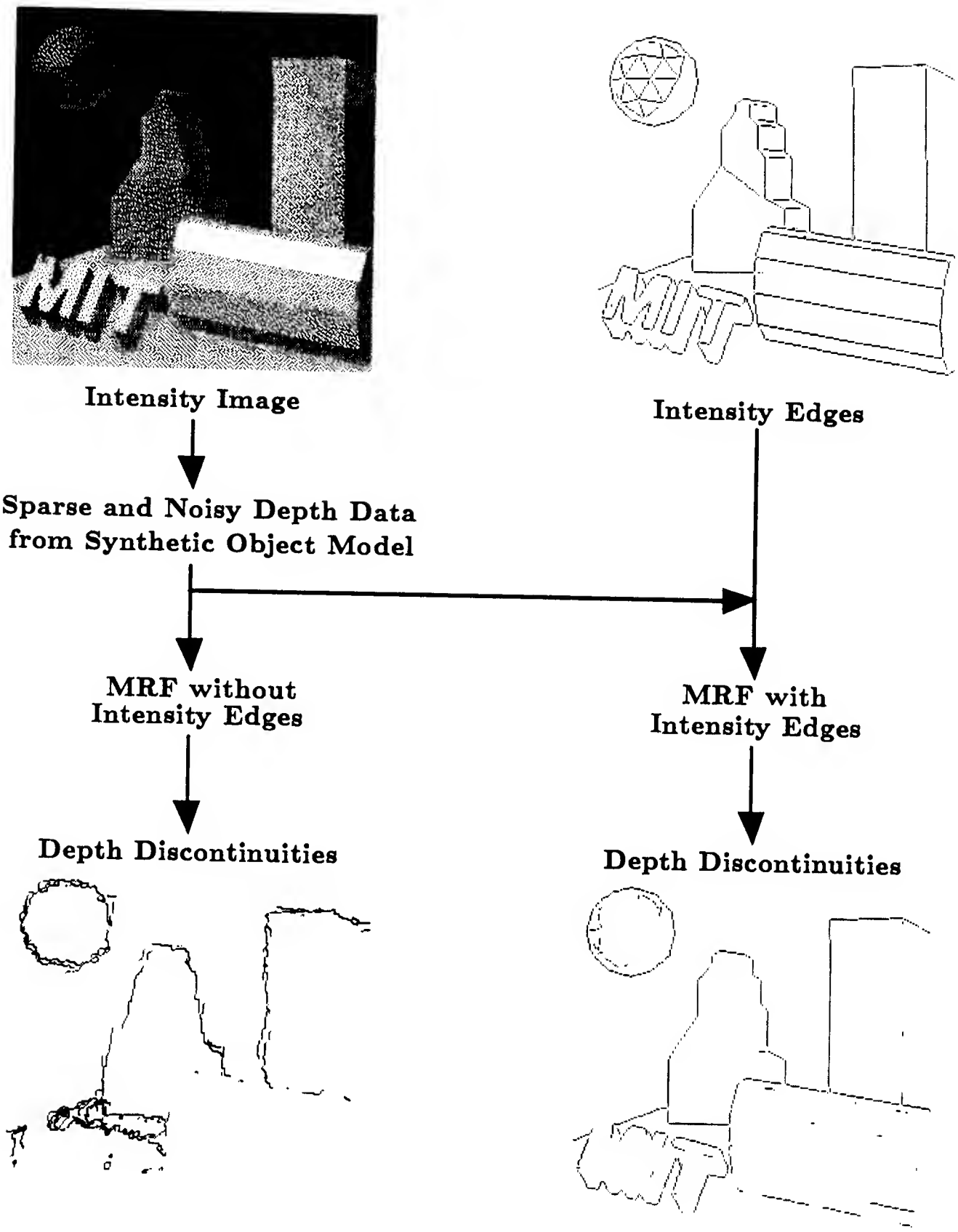

Figure 4: The MRF process and its result on a synthetic image. Almost all depth discontinuities are found when intensity edge coupling is utilized. The steepness of the geodesic dome's boundary leads to false discontinuity identification.

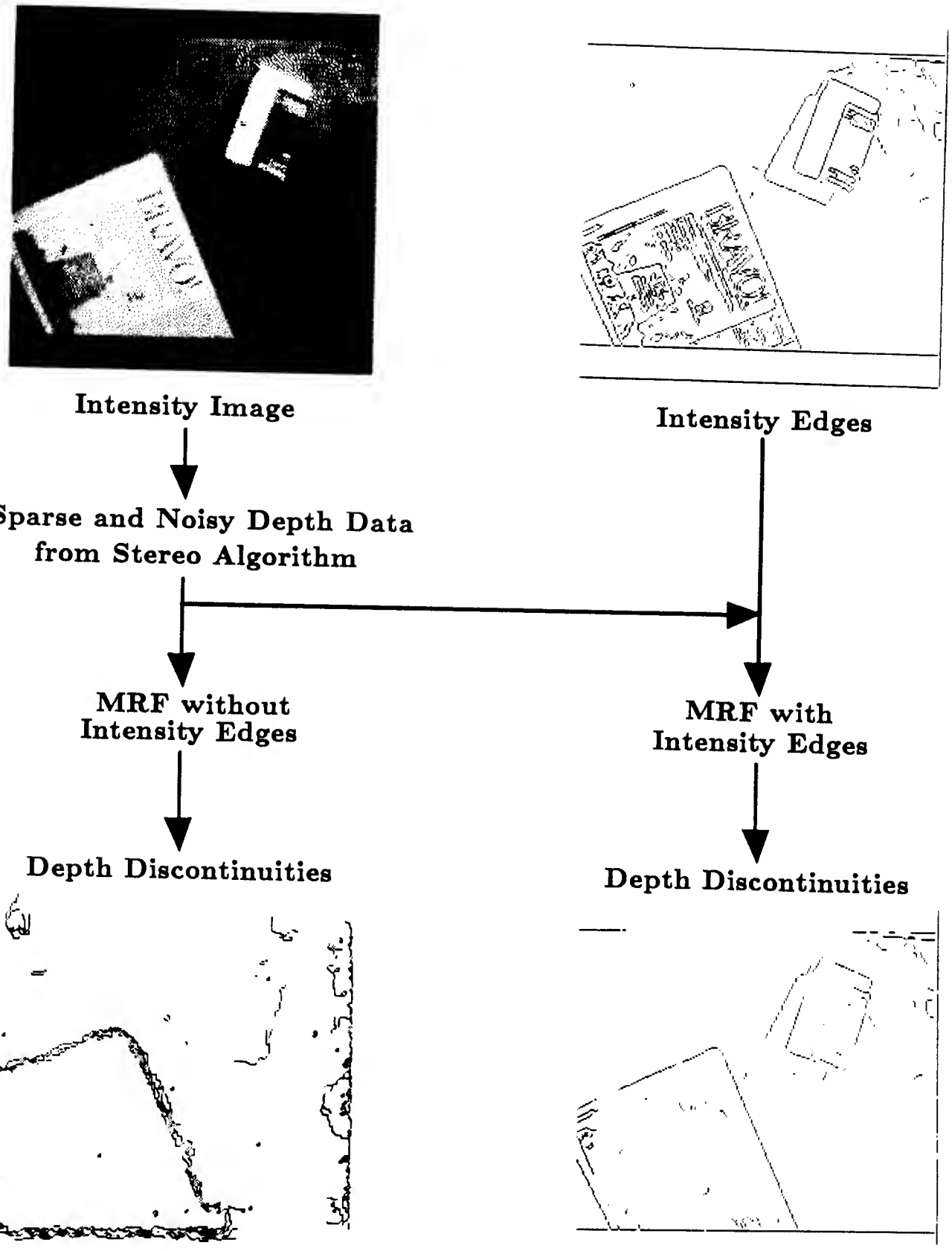

Figure 5: The MRF process and its result on a real image with computed stereo data. For both cases the texture on the newspaper has disappeared; however, without intensity edges, the small box on the upper right also disappears. When intensity edges are used some of the box's borders persist and the newspaper border is well localized.

erywhere except at the intensity edges shown in the upper right image of Figure 4. The external field, e_i^j, equals one only at the intensity edges pixels. The depth discontinuities found are shown on the lower right of Figure 4. Nearly all the intensity edges due to surface orientation and texture are eliminated. In some places, such as near the geodesic sphere's boundary, the surface slope alone is large enough to yield a depth discontinuity.

Another representative image–this time a real image–is shown in Figure 5 where a stereo algorithm produced the sparse depth data. The right image from the stereo pair appears on the upper left of Figure 5. This scene consists of a tall stack of newspapers and a small box or carton. The stereo depth data and the reconstructed surface are not shown. Once again we consider two cases, depending on whether or not the intensity edges are utilized. Without the intensity edges, as with the synthetic stereo results, the depth discontinuities are poorly positioned and ragged. However, with the intensity edges (upper right of Figure 5), the discontinuities on the lower right agree reasonably well with the object boundaries.

For these stereo image results, a few difficulties are worth mentioning. A large depth discontinuity along the top left of the newspaper boundary is not found. The stereo algorithm produced very poor depth data at this location and positioned the depth change roughly 5 pixels above the newspaper intensity edge used by the MRF process. Also the small box's shadow yielded a small disparity that created a depth discontinuity. The box itself also had a small disparity so that modifying MRF parameters to eliminate the shadow discontinuity would have eliminated the box's discontinuity. This sort of variability is inevitable until a reasonable method for local parameter estimation is developed.

Situations can arise wherein discontinuity detection is hampered when the intensity edge sites do not coincide with the sites at which external depth data are provided. Figure 6 displays a possibility where a depth discontinuity should form between features A-1 and A-2 inclusive. However, the discontinuity can only form on the intensity edge at B-1 and, because of depth filling and smoothing, the discontinuity may be *washed out.* The *washing out* depends primarily on the depth difference, the separation between edges A-1 and A-2 and the smoothing parameters. If edge B-1 were on A-1 or A-2, then the discontinuity could form readily. One approach to avoid this coincidence problem is to project a cone of influence about the intensity edge

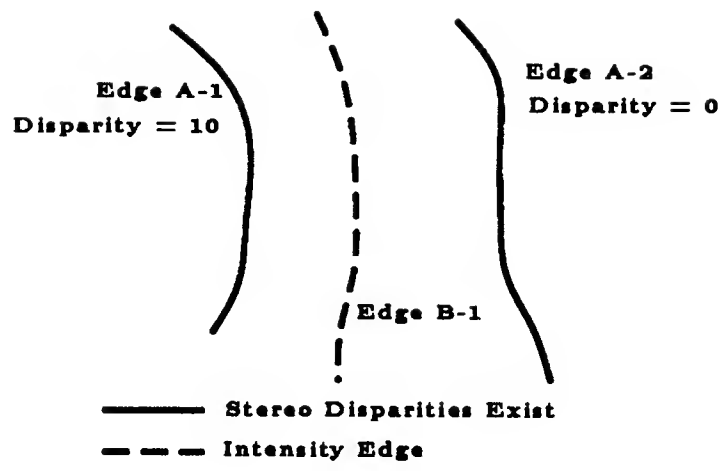

Figure 6: The disparities at edges A-1 and A-2 suggest that a depth discontinuity should be formed somewhere between A-1 and A-2. Yet, because of depth process smoothing, the depth difference at intensity edge B-1 may be too small to support a discontinuity. No discontinuity will form due to this 'misalignment' of edges.

location. Then the discontinuities could form not only at the intensity edges but also for one, two or more pixels on either side of the edge. This has the disadvantage of leading to somewhat poorly localized and ragged edges. Straightness of the resulting line process is enforced locally by the intrinsic prior of the line process when the cone of influence is no larger than the line process neighborhood. Another approach, used here, was to avoid the *washing out* by an appropriate selection of the coupling parameters. More work must be done in this area.

5 Coupling Intensity Edges to Sparse Motion Data

The simplicity of limiting discontinuities to a subset of intensity edges immediately suggests its use for other vision modules. The same principles employed for the stereo depth application have been utilized on motion data. As with depths, motion fields both from synthetic data and a feature-based motion algorithm have been used to identify motion discontinuities and to smooth and fill the sparse motion field. The difference is that motion is a

vector field; depth is not.

The MRF energy of equation 7 is modified by replacing the random field variable, F, by a vector random field, $\vec{M}$. Likewise, the external field, G, becomes a vector field, $\vec{N}$. The MRF energy is:

$$U_i(x|y) = \alpha\gamma_i|\vec{M}_i - \vec{N}_i|^2 + \sum_{j\in nn}(1 - l_i^j)|\vec{M}_i - \vec{M}_j|^2 +$$

$$\sum_{j\in<h,v>}\left[\beta U_C(l_i^j) + \beta'(1 - e_i^j)l_i^j\right] \tag{8}$$

where $\vec{M} = u\hat{e}_x + v\hat{e}_y$ with a similar definition for $\vec{N}$ and where $|\vec{M}_i - \vec{M}_j|^2 = (u_i - u_j)^2 + (v_i - v_j)^2$. The input field $\vec{N}$ contains the two components of the optical flow; the output is $\vec{M}$ or equivalently, (u_i, v_i) for all lattice sites i. With this energy formulation, motion field direction discontinuities are not identified, only magnitude discontinuities are marked.

A specialized motion algorithm, such as Horn and Schunk's[12], can be used to compute the motion field for input to the MRF. The motion data employed here derive from a parallel algorithm[14] that provides match scores much like the previously used stereo algorithm. Match scores provide a local measure of trust for the motion data but are not utilized here. Rather than splitting the problem into early and middle vision parcels, an alternative approach uses the MRF machinery to compute the motion field in addition to segmenting the images[20].

Figure 7 illustrates some results on a simple synthetic motion sequence. The image contains a white square with a small grey texture marking moving diagonally across a grey and black background. The motion field is non-zero only on the white square and its texture marking where both x and y components exist. Roughly 5% of the image motion data is input to the MRF. The bottom half of figure 7 shows the motion discontinuities identified both with and without intensity edge information. Again, the intensity edges significantly enhance the localization of "nice" motion discontinuities.

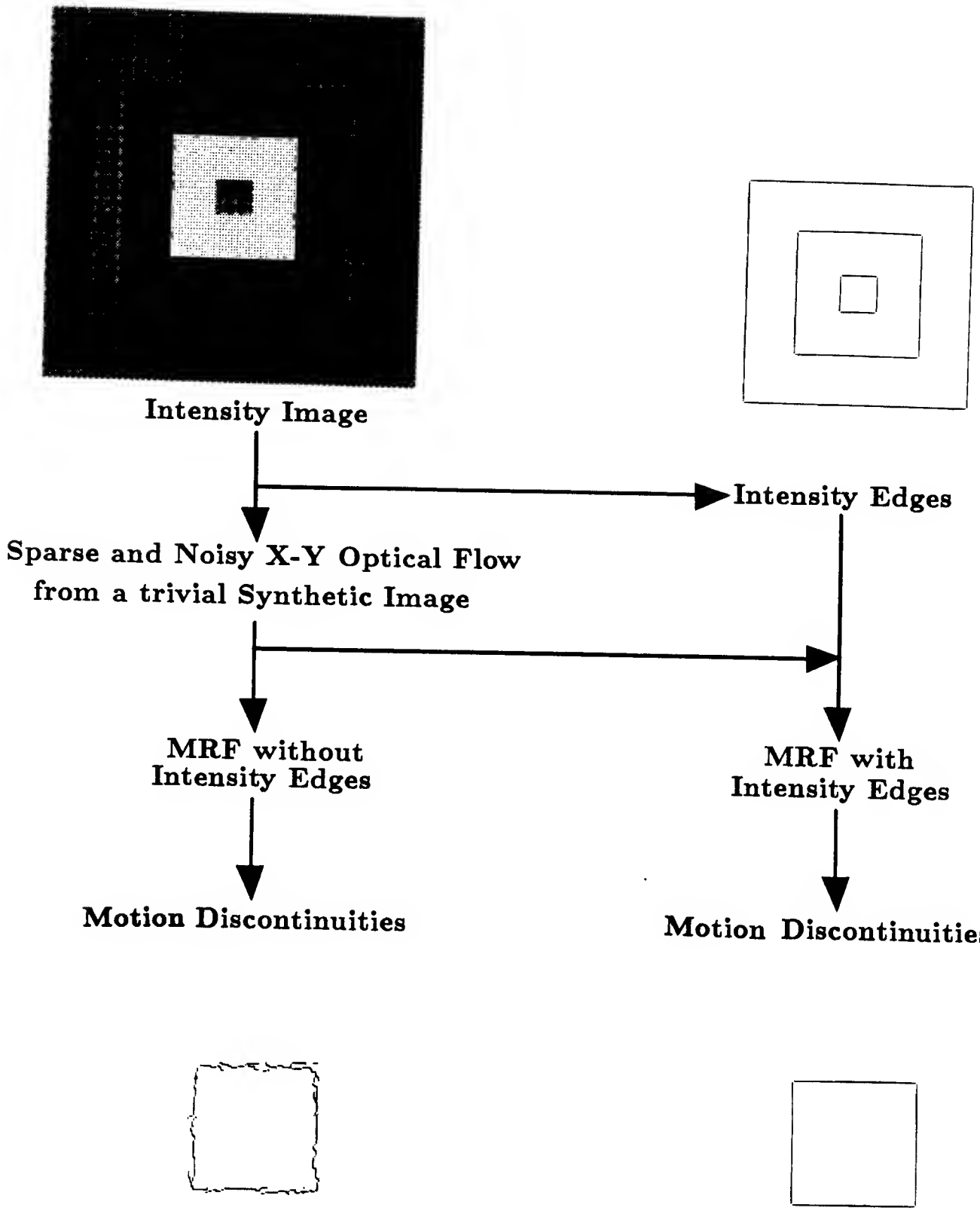

Figure 7: The MRF process and its result on synthetic motion data. Motion data exists at only 5 percent of the image pixels.

6 Discussion

6.1 Central role of intensity edges

The results presented here support the idea that intensity edges can be used as the primary cue to help detect, complete and precisely locate the discontinuities in the other processes such as depth, motion, texture and color. As we mentioned earlier, the reason for this is that discontinuities in depth, surface orientation, motion, texture and color typically originate large gradients in the image intensity, i.e. edges. Texture boundaries, for instance, can be synthesized without any intensity edge; it is sufficient to look around to convince ourselves that in the real world most of the texture boundaries occur together with an intensity edge. The same is true for motion discontinuities. Color boundaries also correspond to brightness boundaries (isoluminant borders exist only in the psychophysics lab!). In addition intensity edges can be better localized than motion, depth, texture and color discontinuities. The case of texture is especially clear: the uncertainty in the location of texture boundaries is no less than the size of the basic elements of texture, called textons[26] and usually several times as much. In most cases stereo cannot provide precise depth discontinuities because of occlusions. Color is in a similar situation because of the coarse scale at which it is computed (the low resolution is imposed by the low signal to noise ratio and the desired insensitivity to small surface markings).

Psychophysics also suggests that intensity information has a privileged role relative to other cues. Cavanagh[5] has shown that only intensity edges can support subjective contours and shadow interpretation. Furthermore, discontinuities portrayed through cues besides intensity edges, are more difficult to see at the level of recognition.

6.2 Open problems in the approach

The preliminary results obtained by integrating intensity edges with depth and motion data are encouraging, as the figures show. There are, however, many open questions that have to be answered before our theory can be regarded as a serious first step towards understanding visual integration. First, there is the question of the overall organization of the integration stage, the nature of the interactions and the couplings between the different

cues. There are also more specific questions about our technique of visual integration and discontinuity detection.

6.2.1 The Structure of Visual Integration

The scheme sketched in figure 8 is a preliminary suggestion for the structure of visual integration. It is close in spirit to the ideas about intrinsic images proposed by Barrow and Tennenbaum[1]. They did not, however, have the powerful theory of coupled MRF models to implement their ideas.

Information about the image intensity has a primary role – intensity edges help the line processes associated with color, texture, motion and depth. Depth itself has also a special role – in a sense, it is the main output of the whole system. Motion, texture and color are coupled to depth. They may not be directly coupled to each other. Notice that *the main couplings are through the line processes*, according to the principles outlined in the introduction. Notice also that local estimates of reliability may be used to control locally the strength of the coupling: we have seen earlier that in the MRF model the coupling between depth and its discontinuities is controlled by the parameter α which is inversely proportional to σ^2.

The line processes may receive data from early algorithms – at this point it is an open question how. In the present implementation the intensity edges are totally driven by external data provided by the Canny edge detector whereas depth and motion do not get external information about discontinuities in depth or motion.

The intensity edges are also coupled with a higher level field that favors configurations of the subjective contour type, providing completion of lines and collinearity on a more global basis than the neighborhood of the line process[22]. The depth line process is coupled with another high-level field that provides the correct constraints on the interactions between contours of overlapping objects. A **T** junction is a clue to occlusion by one of the two surfaces bounded by it; an **X** intersection indicates that one of the surfaces may be transparent. The high-level features couple these configurations of the line process to the appropriate states of the depth process. If no values are locally available, default values for *in front* and *behind* are given to the depth process.

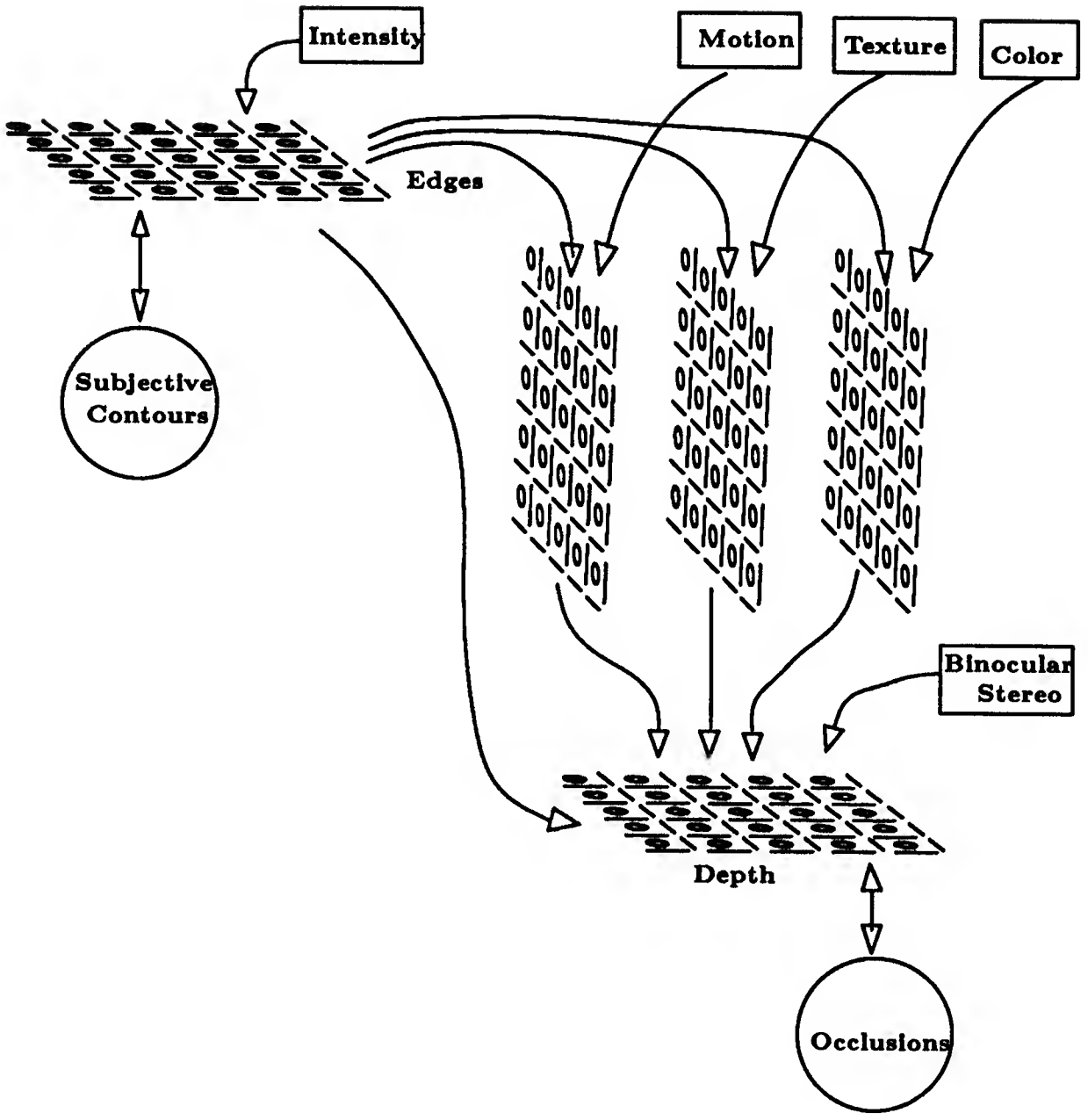

Figure 8: The organization of the integration stage. Each of the processes is coupled to its line process. Intensity data feed into the motion, color, texture and depth line processes. The line processes are not hidden processes: they may also receive data from specialized discontinuity detectors. The intensity line process gets input data from Canny edges. It is coupled to a higher level field which implements constraints of line continuation and collinearity on a more global basis than the neighborhood system of the line process. The line process associated with the depth process is also coupled to a higher level field which implements the appropriate constraints underlying occlusions of surfaces. The plausibility of interactions between motion, texture and color is an open question.

6.2.2 Detailed Questions

Other open questions are: *integration of additional visual cues*, *local vs. global constraints on the line process*, *tolerance in registration*, *multiresolution fields*, *approximative algorithms and neural implementations* and *learning of parameters from examples.*

Integration of additional visual cues As figure 8 shows, we plan to integrate other visual cues with stereo, motion and intensity data. In particular, we will include texture and color. Because texture boundaries usually depend on changes of material or sharp changes in surface orientation, they could be used to support the line processes in the depth and motion modules. For color the goal is to find boundaries that delineate regions of constant albedo (at a coarse resolution, since small surface markings should not be "seen" at this stage). As in the case of depth and motion, intensity edges play a critical role for these two additional visual modules. Hurlbert and Poggio (see [21]) have sketched a possible scheme for coupling albedo with intensity edges.

It is important to notice that the combination of several visual cues not only allows reinforcement of evidence for, say, a depth discontinuity, but also achieves a classification of an intensity edge in terms of its underlying physical cause: for instance, whether it is due to a shadow or a depth discontinuity. Clearly, psychophysics can give useful indications of which interactions are important in the human visual system.

Local versus global constraints on the line process The line process provides a means for imposing important physical constraints on the discontinuities such as: continuity, relative spatial isolation and possibly collinearity. These constraints are enforced by using appropriate cliques and associated *energy* values. However, in our experience with Markov Random Field models applied to real data, a problem has emerged with the use of the line process. In many cases the property of collinearity that can be enforced in this way remains too local: discontinuities tend to be too jagged and sometimes even broken when integration with intensity edges is *not* used. How can one enforce the property of continuity or simply collinearity over larger distances within the MRF framework? The basic idea that we have begun to explore is to have a higher–level MRF that consists of "features", such as

straight lines of different orientations, with its prior probability distribution, coupled (bidirectionally) with the line process lattice (see figure 8).

Tolerance in registration When data from different cues are combined, say from intensity and from stereo, they must be registered. Spatial coincidence is the main constraint exploited here. In general, however, one cannot expect that discontinuities in depth and intensity will always have *exactly* the same location. Because of errors in the early vision processes, effects of filtering, photometric effects and so on, depth discontinuities may be offset by one or more pixels from intensity edges. To deal with this registration problem the cone of influence might be useful, in which the intensity edges facilitate (or don't veto) the formation of depth discontinuities. The cone of influence size should be on the order of the line process neighborhood. In this way the line process constraints will ensure collinearity within the cone-of-influence. Again, important information will come from psychophysics: we expect to learn how alignment of, for instance, intensity edges with depth discontinuities affects human vision.

Learning parameters from examples A critical problem in using MRFs is the problem of parameter estimation. The performance of the scheme depends critically on the natural temperature of the field, the potentials associated with the clique configurations, the coupling between the lattices, and so on. Parameter estimation should provide estimates for these factors; possibly by learning from a set of examples.

Does integration influence early vision modules? In our computational approach to integration we have tacitly assumed that information flows from the early vision modules to the integration stage — the coupled MRF system — but not backwards. The output of say, stereo, is modified by the outputs of other modules at the level of the MRFs but the stereo process itself — the matching, for instance — is not affected. The decision to neglect feedback interactions, from the integration stage to the early processes, in the present version of our theory is mainly due to reasons of simplicity. Without modifying our scheme in an essential way, it is easy to incorporate backward effects from the integration stage by assuming that the whole process from early vision algorithms to the integration stage can be controlled by a higher–

order system taking into account higher–level goals and the available results. If recognition is the goal, for instance, the current results of the recognition operation on the integrated information can control which early processes to apply, where, and how (i.e. which parameters to use). In this case, one may hope to develop a useful theory of integration without worrying at first about the problem of feedback.

A different possibility is that interactions between the integration stage and the early vision modules are an essential part of any integration theory and cannot be neglected even in a first-order approximation. In an extreme case one might not be able to separate the integration stage usefully from the early vision modules and even the modules one from another.

In principle, this is possible. The algorithms for the early processes can be regarded in several cases as MRFs themselves (regularization algorithms are special cases of MRFs[2,23]). Thus our coupling schemes for integration can be extended to couple the early processes. In practice, we expect that parameter estimation may become a very serious problem once the early vision processes are tightly coupled.

Hardware implementations As discussed elsewhere[19,21] the coupled MRF models used here can be implemented efficiently in mixed digital and analog hybrid networks. It is interesting that, the interaction underlying coupling between fields is of the type of a multiplication, logical-and or veto operation. These operations have some intriguing possible implementations in terms of the properties of synapses.

While it is certainly possible to implement the same mixed deterministic and stochastic algorithms described here in, say, VLSI technologies, it is also interesting to explore approximative deterministic algorithms that may be simpler and more efficient. Marroquin[16] has provided an encouraging initial analysis along with estimates of convergence properties.

References

[1] H. Barrow and M. Tenenbaum. *Recovering Intrinsic Scene Characteristics from Images.* Academic Press, New York, 1978.

[2] M. Bertero, T. Poggio, and V. Torre. *Ill-Posed Problems in Early Vision.* A.I. Memo No. 924, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1986.

[3] Andrew Blake and Andrew Zisserman. *Visual Reconstruction.* The M.I.T. Press, Cambridge, MA, 1987.

[4] John F. Canny. *Finding Lines and Edges in Images.* Technical Report TM-720, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1983.

[5] Patrick Cavanagh. Reconstructing the third dimension: interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing*, 37:171–195, 1987.

[6] Haluk Derin and Howard Elliott. Modeling and segmentation of noisy and textured images using gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(1):39–55, January 1987.

[7] Michael Drumheller. Connection Machine stereo matching. In *Proceedings of the AAAI*, pages 748–753, August 1986.

[8] Michael Drumheller and Tomaso Poggio. On parallel stereo. In *Proceedings of the IEEE Conference on Robotics and Automation*, San Francisco, CA, 1986.

[9] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[10] W. Eric L. Grimson. *From Images to Surfaces.* M.I.T. Press, Cambridge, MA, 1981.

[11] W. Danny Hillis. *The Connection Machine.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1985.

[12] B. K. P. Horn and B. G. Schunk. Determining optical flow. *Artificial Intelligence*, 17, 1981.

[13] James J. Little, Guy E. Blelloch, and Todd Cass. Parallel algorithms for computer vision on the Connection Machine. In *Proceedings of the International Conference on Computer Vision*, London, England, June 1987.

[14] James J. Little, Heinrich Bulthoff, and Tomaso Poggio. Parallel optical flow computation. In *Proceedings of the Image Understanding Workshop*, pages 417–431, Los Angeles, CA, February 1987.

[15] David Marr and Tomaso Poggio. A theory of human stereo vision. *Proc. R. Scoc. London B*, 204:301–328, 1979.

[16] Jose L. Marroquin. Deterministic bayesian estimation of markovian random fields with applications to computational vision. In *Proceedings of the International Conference on Computer Vision*, London, England, June 1987.

[17] Jose L. Marroquin. *Probabilistic Solution of Inverse Problems.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1985.

[18] Jose L. Marroquin. *Surface Reconstruction Preserving Discontinuities.* A.I. Memo No. 792, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, August 1984.

[19] Jose L. Marroquin, Sanjoy Mitter, and Tomaso Poggio. Probabilistic solution of ill-posed problems in computational vision. *J. Amer. Stat. Assoc.*, 82:76–89, 1987.

[20] David W. Murray and Bernard F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(2):220–228, March 1987.

[21] Tomaso Poggio. *Integrating Vision Modules with Coupled MRF's.* Working Paper No. 285, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1985.

[22] Tomaso Poggio and Edward B. Gamble Jr. In preperation.

[23] Tomaso Poggio, Harry Voorhees, and Alan Yuille. *Regularized Solution to Edge Detection.* A.I. Memo No. 833, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, May 1985.

[24] Demetri Terzopoulos. *Computing Visible Surface Representations.* A.I. Memo No. 800, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, March 1985.

[25] Demetri Terzopoulos. The role of constraints and discontinuities in visible-surface reconstruction. In *Proceedings of the IJCAI*, pages 1073–1077, August 1983.

[26] Harry Voorhees and T. Poggio. Detecting textons and texture boundaries in natural images. In *Proceedings of the International Conference on Computer Vision*, London, England, June 1987.